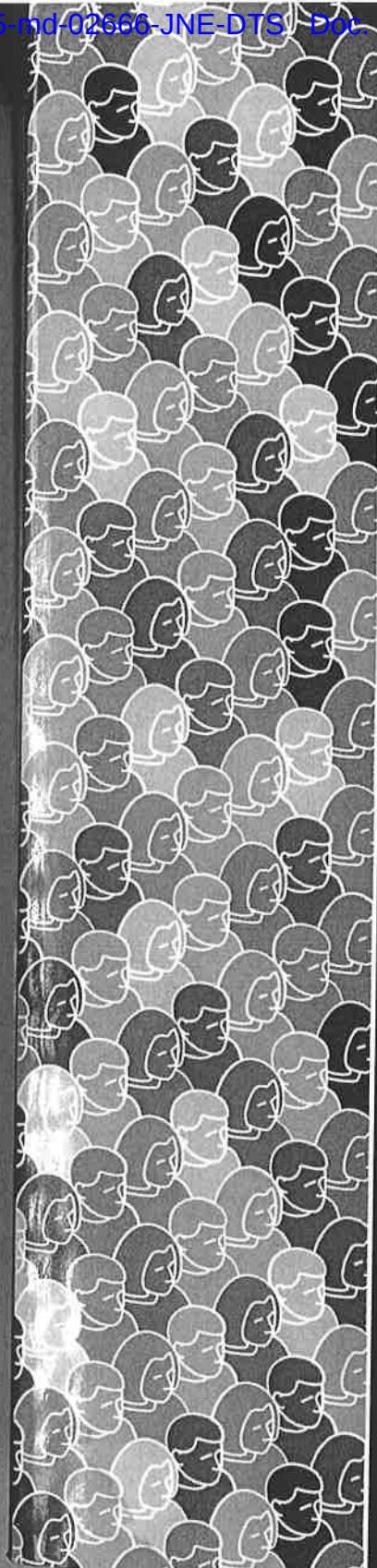


# EXHIBIT 6



# **Essentials of Epidemiology in Public Health**

**Ann Aschengrau  
George R. Seage III**

# 8

## Cohort Studies

### LEARNING OBJECTIVES

By the end of this chapter, the reader will be able to:

- Distinguish between the various types of cohort studies, including open, fixed, closed, retrospective, prospective, and ambidirectional designs.
- Describe the key features of conducting cohort studies, including the selection of the exposed and unexposed populations; the sources of information on the exposure, outcomes, and other key variables; approaches to follow-up; calculating person-time; and data analysis.
- Discuss the strengths and limitations of cohort studies.

### Introduction

Although experimental studies are scientifically rigorous, they are often infeasible because of difficulties enrolling participants, high costs, and thorny ethical issues. For example, an experimental study of a substance that increases a person's risk of developing a disease is infeasible because it is unethical to give a potentially harmful substance to a healthy person. Consequently, epidemiologic research consists mainly of observational studies. These studies are considered "natural" experiments, because the investigator acts as a disinterested observer merely letting nature take its course. Observational studies take advantage of the fact that people expose themselves to noxious or healthy substances through their personal habits (such as by drinking alcoholic beverages or taking vitamins), choice of occupation (such as asbestos insulation worker), place of residence (for example, near a lead smelter), and so on. Consequently, many issues that

make it infeasible to conduct an experimental study—particularly those related to ethics—are mitigated in an observational study. For example, although it is unethical to conduct an experimental study of the impact of cigarette smoke on the developing fetus by randomly assigning newly pregnant women to either a smoking or nonsmoking group, it is perfectly ethical to conduct an observational study by comparing women who choose to smoke during pregnancy with those who choose not to do so.

As described in Chapter 6, the cohort study is one of the two principal types of observational studies. (The other type, the case-control study, is covered in Chapter 9.) A classical cohort study examines one or more health effects of a single exposure. Subjects are defined according to their exposure status and followed over time to determine the incidence of the health outcomes. This chapter describes the design, conduct, and analysis of cohort studies. First, cohort study terms are defined, and then a detailed discussion of each aspect of these studies is provided. Many design features of cohort studies emulate those of well-done experimental studies, because the goal of both types of studies is to obtain high-quality results.

## Cohort Study Definitions and Overview

The term *cohort* comes from the Latin word *cohors*, meaning a group of soldiers.<sup>1(p359)</sup> In ancient Rome, a cohort was one of 10 divisions of a legion, the major unit of the Roman army. Once a cohort was formed, no new soldiers were added, and so soldiers remained in the same cohort for the duration of their service. Attrition occurred mainly through death.

Today, we use the word cohort to characterize “any designated group of persons who are followed or traced over a period of time.”<sup>2(p31)</sup> The term is also used to describe a group of individuals with a common characteristic or experience. For example, a birth cohort consists of individuals who are born during a particular year or period, and so share many similar experiences (such as “baby boomers”). Although *cohort study* is the standard term used to describe an epidemiologic investigation that follows groups with common characteristics, some investigators use the terms *follow-up*, *incidence*, and *longitudinal study*.

Several additional terms are used to describe cohort studies. These terms are related to: the characteristics of the population from which the cohort is derived, whether the exposure changes over time, and whether there are losses to follow-up (see Table 8-1).

### Types of Population Studied

The first type of cohort is conducted in an *open* or *dynamic* population. Individuals in an open population may enter or leave at any time because its membership is defined by a changeable characteristic such as smoking

TABLE 8-1 Characteristics of Cohort Studies

Type of population studied	Defined by	Follow-up	Appropriate measure of disease frequency
Open or dynamic	Changeable characteristic	Members come and go; losses may occur	Incidence rate
Fixed	Irrevocable event	Does not gain members; losses may occur	Incidence rate
Closed	Irrevocable event	Does not gain members; no losses occur.	Cumulative incidence

cigarettes, drinking alcohol, having a certain occupation, or living in a specific geographic area. For example, a person is a member of the open population of New York City residents only as long as he or she lives in New York City. Cohort studies that are conducted in an open population usually take into account population changes such as in- and outmigration, and so the incidence rate is the most suitable measure of disease frequency for monitoring health outcomes in this setting. A cohort study of cancer incidence was recently conducted among never-married men who were aged 25 to 54 years and resided in San Francisco, California, from 1973 to 1990.<sup>3</sup> These eligibility criteria were selected in order to identify a population with a high prevalence of HIV infection. This group is considered an open population because it is defined by the changeable characteristics of marital status, age, and place of residence. If a man got married, aged beyond 54 years, or moved away from San Francisco, his membership in the population ended and he was no longer eligible for the study. The study investigators monitored these eligibility changes among the men enrolled in the study, and an estimated 1,390,000 person-years of observation was accrued during the follow-up period (see Chapter 2 for more details about person-time calculations). The person-time data were used as the denominators for the incidence rates of non-Hodgkin's lymphoma and other cancers. The study found that the incidence of non-Hodgkin's lymphoma increased 20-fold from 1973 to 1990, while that of anal cancer did not increase. These findings suggest that the causal mechanisms of HIV-related cancer differ for specific cancers.

Cohorts may also be formed on the basis of inalterable characteristics (see Table 8-1). A *fixed cohort* is defined by an irrevocable event—for example, undergoing a medical procedure, giving birth to a child, serving in the military, eating contaminated food at a picnic, or being present at a man-made or natural disaster. Thus, exposures in a fixed cohort do not change over time. In addition, the groups are followed from a defined starting

point (usually marked by the event) to a defined ending point. The World War II atomic bomb survivors from Hiroshima and Nagasaki, Japan, comprise one of the best known fixed cohorts to be studied for biological effects of acute radiation exposure.<sup>4</sup> For nearly half a century, researchers have monitored mortality and cancer incidence rates among approximately 94,000 exposed residents who survived the bombings and 27,000 unexposed residents who were outside the city when the bombs were dropped. The incidence rates of various outcomes have been periodically compared between these two groups during the 50-year follow-up period. Incidence rates are the appropriate measure of disease frequency in a fixed cohort when the population experiences losses to follow-up. (Loss to follow-up means that investigators are unable to trace all members of a cohort to determine if they became ill or died.)

The third type of cohort is the *closed cohort*. Like the fixed cohort, a closed cohort is defined by an irrevocable event and has defined starting and ending points for follow-up. The difference between the two is that a closed cohort has no losses to follow-up. For example, a closed cohort study might be conducted among people who attended a party to determine if eating certain foods increased the risk of gastroenteritis during the week following the party. Thus, everyone who attended the party is a member of the closed population who is eligible for the study. Follow-up would start at the end of the party (assuming that all of the contaminated food was eaten by then, and that the illness could occur immediately after ingesting the contaminated food), and it would end 7 days later. No members of the population would be lost, because the observation period is short. Cumulative incidence or average risk is typically used as the measure of disease frequency in this setting, because there are no losses to follow-up.

### Characterization of Exposure

Regardless of what type of cohort study is conducted, participants are grouped according to their exposure and followed over time in order to compare the incidence of symptoms, disease, or death. Usually two groups are compared, such as an exposed and unexposed group. The exposed group is called the *index group*, and the unexposed group is termed the *referent* or *comparison group*. It is necessary for the investigator to specify a minimum amount of exposure to qualify for the exposed group. For example, studies of cigarette smoking may use a lifetime history of smoking at least 100 cigarettes to qualify for membership in the exposed smoking group. In addition, when investigators are unable to find truly unexposed people to serve in the comparison group, people with low exposure are selected.

It can be difficult to classify the exposure status of subjects, particularly when the exposure changes over time. For example, a woman might

begin smoking as a young adult, smoke cigarettes for a few years, stop while she is pregnant, and then return to smoking after giving birth. Thus, depending on the period, this woman could be considered a nonsmoker, smoker, or ex-smoker. The investigator might decide to consider a woman exposed if she ever smoked cigarettes at any time in the past. However, such a simplistic exposure definition might miss an association that is confined to certain aspects of the exposure. For example, the association may only be present for high-intensity smoking (for example, more than two packs per day) or long-duration smoking (for example, more than 10 years).

Thus, whenever possible, investigators divide the exposed group into levels (for example, high, medium, and low exposure) enabling investigators to assess the presence of a dose-response relationship. A dose-response relationship means that the risk of disease increases as the intensity or duration of exposure increases. This categorization can be complicated, because there are many ways to characterize an exposure level. For example, one could use the maximum level ever experienced by the individual, the cumulative level as of a certain date, or the average exposure level over a time period. Exposure to cigarette smoke is often characterized by the number of pack-years that the person has accumulated. This composite measure is calculated by multiplying the average number of packs smoked per day by the number of years smoked at that intensity. Thus, a person who has smoked two packs a day for 20 years has accumulated 40 pack-years of exposure.

### **Follow-up and Outcome Assessment**

During the follow-up period, the exposed and unexposed groups are monitored for the outcomes under study. As in experimental studies, more than one outcome is typically investigated. The outcomes of interest depend on the research question and may include precursors or first occurrence of disease, disease recurrence, or death. For example, a cohort study of U.S. pulp and paper mill workers examined all causes of death.<sup>5</sup> As another example, a cohort study of Canadian farm couples examined their fecundability, that is the monthly probability of becoming pregnant.<sup>6</sup>

At the start of follow-up, members of the cohort are at risk for but have not yet experienced the outcome(s) of interest. By the end of the follow-up period, a proportion of the cohort (up to 100%) will have experienced the outcome(s) under study. The length of the follow-up period can range from a few hours for infectious diseases to several decades for diseases such as cancer or cardiovascular disease. The longer the follow-up period, the more difficult it is to trace and maintain contact with study subjects and the more expensive the study. Although follow-up has become even

## 198 ESSENTIALS OF EPIDEMIOLOGY IN PUBLIC HEALTH

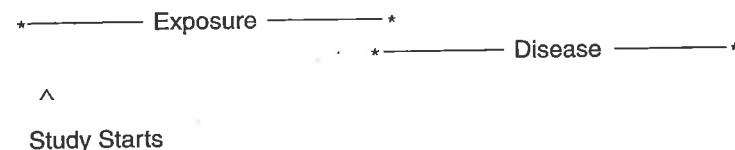
more difficult in recent years because of increased population mobility, newly available resources such as those on the Internet have helped to offset these difficulties. Follow-up rates should be as high as possible in order to ensure the validity of the study. If the health outcomes of those who are lost to follow-up are different from those who are not lost, the study results may be biased. There is no magical follow-up rate that guarantees the validity of a study (other than 100%), but most epidemiologists are satisfied with follow-up rates higher than 90%.

### Timing of Cohort Studies

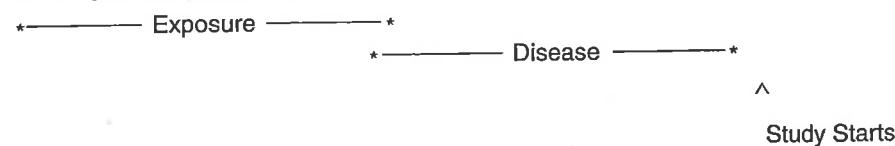
Three terms are used to describe the timing of events in a cohort study: (1) *prospective*, meaning to look forward in time; (2) *retrospective*, meaning to look back in time; and (3) *ambidirectional*, meaning to look both ways (see Figure 8-1). In order to avoid confusion with other types of studies, the terms retrospective, prospective, and ambidirectional should always modify the word *cohort* and never be used alone.

In a *prospective cohort study*, participants are grouped on the basis of past or current exposure and are followed into the future in order to

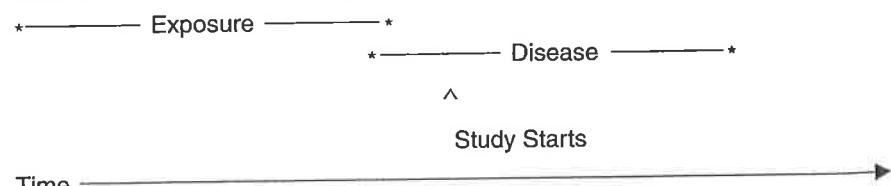
#### Prospective Cohort Study



#### Retrospective Cohort Study



#### Ambidirectional Cohort Study



**FIGURE 8-1.** Timing of Cohort Studies

observe the outcome(s) of interest. When the study commences, the outcomes have not yet developed and the investigator must wait for them to occur. The timing of prospective cohort studies is similar to that of experimental studies. For example, a prospective cohort study was conducted to investigate the effect of low-level lead exposure on intelligence and academic achievement among children.<sup>7</sup> Investigators enrolled children at birth and followed them until they were 10 years old. Investigators obtained the children's blood lead levels at birth and several times thereafter. When the children reached 10 years of age, investigators conducted development assessments to determine if prior lead exposure was associated with IQ and school performance.

In a *retrospective cohort study*, both the exposures and outcomes have already occurred by the time the study begins. This type of investigation studies only prior outcomes and not future ones. In fact, retrospective cohort studies often evaluate diseases that occurred many years ago. For example, in the 1990s, Swedish investigators undertook a nationwide retrospective cohort study to evaluate the association between breast implants and connective tissue disorders.<sup>8</sup> The study included more than 7000 women who had undergone breast implant surgery from 1964 through 1993 for cosmetic reasons or for reconstruction after breast cancer. Investigators collected data on the occurrence of connective tissue disorders following the implant surgery. Thus, the surgery and all of the outcomes under study had occurred by the time the investigators began the study. The researchers found no association between breast implants and connective tissue disease.

An *ambidirectional cohort study* has both prospective and retrospective components. For example, the Air Force Health Study is an ambidirectional study of the men who were involved in aerial spraying of herbicides (including the dioxin-contaminated Agent Orange) during the Vietnam War.<sup>9</sup> The purpose of the Ranch Hand Study is to determine if these men have an increased risk of adverse health and reproductive outcomes. The retrospective portion of the study conducted analyses of cancer and mortality that occurred from the men's first tour of duty in Vietnam through the 1980s.<sup>10,11</sup> The prospective component will monitor the health of these men for many years to come.

How does an epidemiologist decide whether to conduct a prospective, retrospective, or ambidirectional cohort study? The answer often depends on the research question, the practical constraints of time and money, and the availability of suitable study populations and records. In addition, the decision usually takes into account the complementary advantages and disadvantages of retrospective and prospective cohort studies. For example, retrospective cohort studies are more efficient (they take less time and money) than prospective studies for investigating diseases that take a long

time to develop and come to diagnosis. This is because the disease and exposure have already occurred.

Minimal information is usually available on the exposure, outcome, and other key variables in retrospective cohort studies, because these studies typically rely on existing records that were not designed for research purposes. For example, a retrospective cohort study of men who worked in a chemical factory in the 1950s might have to depend on employment records to identify exposed and unexposed cohort members. These records may only contain the person's name, date of birth, department, job title, and years of employment. A detailed picture of that person's actual job duties and chemical exposures and information on important habits such as cigarette smoking may be missing. In addition, follow-up and ascertainment of the outcomes may be hampered because the historical records may not have information to facilitate locating the study subjects many years later.

In contrast, investigators in a prospective cohort study can usually obtain more detailed information on exposures and other key variables because they have more control of the data collection process and can gather information directly from the participants. Follow-up may also be easier, because the investigator can obtain tracing information from participants and can maintain periodic contact with subjects. In addition, prospective cohort studies are considered less vulnerable to bias because the outcomes have not occurred when the cohort is assembled and the exposures are assessed.

## **Issues in the Selection of Cohort Study Populations**

### **Emulation of Experimental Studies**

The investigator in an experimental study assigns participants to the experimental or comparison group. When the assignment is made randomly and the sample size is sufficiently large, these groups will have nearly identical distributions of baseline characteristics. In contrast, participants in cohort studies enter the exposed and unexposed groups usually on the basis of self-determined behaviors and events. Consequently, the results of cohort studies may be difficult to interpret because of differences between the exposed and unexposed groups that influence the risk of developing the disease. Consider, for example, a cohort study of the effect of maternal cigarette smoking during pregnancy on the risk of birth defects among the offspring. Women who smoke cigarettes during pregnancy are more likely to consume alcoholic beverages than women who do not smoke.<sup>12</sup> Because heavy consumption of alcoholic bev-

erages increases the risk of having a baby with congenital anomalies regardless of whether a woman smokes cigarettes,<sup>13</sup> an association between cigarette smoking and birth defects might really be due to the alcohol consumption.

Epidemiologists try to minimize these types of problems by translating the key features of experimental studies to cohort studies. First, even though it is not possible to randomly assign exposures in cohort studies, differences between groups can be minimized by carefully selecting the exposed and unexposed groups to be as similar as possible. In fact, the ideal comparison group would consist of exactly the same individuals had they not been exposed. For example, the Air Force Health Study compared an exposed group of 1300 Air Force servicemen who flew herbicide spraying missions during the Vietnam War with an unexposed group of Air Force servicemen who flew cargo missions in Southeast Asia during the war that did not involve handling or spraying herbicides.<sup>10</sup> The baseline demographic characteristics and health status of the exposed and unexposed groups were quite similar, because all the men underwent the same selection procedures for serving in the Air Force and flying missions during the Vietnam War.

A second key feature of experimental studies that is translated in cohort studies is the use of placebos. Recall that placebos are used in experimental studies to match as closely as possible the experiences of the experimental and comparison groups and to permit masking of the study investigators (among others). Although cohort studies cannot use placebos, they can match the experiences of the exposed and unexposed groups by selecting individuals with similar characteristics. Investigators sometimes enhance this process by using special matching criteria when selecting unexposed subjects. For example, if investigators want to ensure that the groups are comparable with respect to age, gender, and race, they would employ a technique known as individual matching. For example, they would select an unexposed 50-year-old white male for every exposed 50-year-old white male in the study. Even without placebos, it is relatively easy to mask investigators to the subjects' exposure status during follow-up and outcome ascertainment in a cohort study. One merely keeps this information from the study staff involved in these tasks. Masking guarantees comparable follow-up efforts and outcome ascertainment, which—in turn—help ensure valid results.

The third key feature of experimental studies emulated by cohort studies is the method for determining the overall sample size as well as the relative size of the exposed and unexposed groups. Like experimental studies, cohort studies must include an adequate number of individuals in order to have sufficient statistical power to detect an association if it truly exists. Investigators determine the sample size by taking into account the

## 202 ESSENTIALS OF EPIDEMIOLOGY IN PUBLIC HEALTH

**TABLE 8-2** Similarities and Differences Between Experimental and Cohort Studies*Similarities*

- Both make comparisons across two or more exposure groups.
- Both follow participants to monitor outcome rates.
- Both usually monitor more than one outcome.
- Both select groups to achieve comparability and efficiency.
- Relative proportion of subjects in compared groups do not reflect that of the general population.

*Differences*

- Experimental study investigators allocate exposure to exposure groups. Cohort study participants choose exposures themselves.
- Experimental studies can use randomization to achieve baseline comparability. Cohort study investigators must carefully select groups to achieve comparability.
- Experimental studies can use placebos to match the groups' experiences and permit masking. Cohort study investigators carefully select groups to match as closely as possible. Cohort studies also use masking.
- Experimental studies are prospective. Cohort studies may be prospective, retrospective, or ambidirectional.

anticipated difference between compared groups, the background rate of the outcome, and the probability of making statistical errors.<sup>14</sup>(pp142-146) The relative sizes of the exposed and unexposed groups are chosen to maximize the power of the study. They are not intended to mimic the frequencies in the general population. For example, an exposure may occur rarely in the general population (for example, 1 exposure per 1000 population), but the frequency in a study population may be 50% (for example, 1000 exposed and 1000 unexposed individuals). The similarities and differences between experimental and cohort studies described so far are summarized in Table 8-2.

### Selection of the Exposed Population

The choice of the exposed group in a cohort study depends on the hypothesis under study, the exposure frequency, and feasibility considerations such as the availability of records and ease of follow-up. The two main types of cohorts—special cohorts and general cohorts—are distinguished by the exposure frequency. *Special cohorts* are assembled to study the health effects of rare exposures such as uncommon occupational chemicals, unusual diets or lifestyles, natural or man-made disasters, or medical procedures. *General cohorts* are typically assembled for common exposures

such as use of oral contraceptives, dietary factors such as vitamin use, and habits such as cigarette smoking and alcohol consumption. General cohorts are often selected from professional groups such as doctors and nurses, volunteers, or residents in well-defined geographic areas in order to facilitate follow-up and accurate ascertainment of the outcome(s) under study.

When assembling a special cohort, epidemiologists must go to the settings where the exposure has, is, or will occur in order to identify the exposed population. For example, workplace sites such as factories and organizations such as unions frequently serve as sources of individuals for occupational cohort studies. Because workers often have more intense and sustained exposures than the general population, cohort studies of occupational groups are an important component of identifying the causes of disease. For example, an occupational cohort study was conducted among 21,863 male and female workers from 12 countries whose jobs involved producing or spraying phenoxy herbicides and chlorophenols to determine if these workers had an increased risk of cancer deaths.<sup>15</sup> Epidemiologists identified subjects for the cohort study by reviewing employment and other records in businesses involved in manufacturing and spraying these chemicals.

Epidemiologists have conducted special cohort studies among individuals undergoing a medical procedure or treatment involving potentially noxious exposures. For example, important data on the risk of cancer following radiation have come from cohort studies of patients with ankylosing spondylitis (a spinal disorder), who were treated with radiation therapy.<sup>16</sup> Records of 87 radiotherapy centers in Great Britain and Northern Ireland were reviewed to identify more than 14,000 patients treated from 1935 to 1954. Investigators found that ankylosing spondylitis patients were more likely than expected to have leukemia and cancers of the lung, bone, and esophagus.

Epidemiologists have conducted cohort studies among groups with unusual diets or lifestyles such as Seventh Day Adventists and Mormons. For example, Seventh Day Adventists do not consume tobacco, alcohol, or pork, and about 50% follow a lacto-ovovegetarian diet. The Adventist Study has monitored the morbidity and mortality experience of approximately 34,000 Seventh Day Adventists in California for several years.<sup>17</sup> The study has found, for example, that male Adventists have lower overall cancer incidence but a higher incidence of prostate cancer than the general population.

As stated previously, epidemiologists assemble a general cohort from professional groups or residents in well-defined geographic areas in order to study common exposures. One of the best known cohort studies based upon the general population is the Framingham Study. Begun in 1948 and

still continuing today, its main purpose is to determine the causes of cardiovascular disease.<sup>18(pp14-29)</sup> The risk factors that have been studied include high blood pressure, high blood cholesterol, physical inactivity, obesity, and smoking—all fairly common exposures. Investigators initially enrolled about 5000 healthy adult residents in Framingham, Massachusetts, a town located about 18 miles west of Boston. In 1948, Framingham was a self-contained community of about 28,000 residents who obtained their medical care from local physicians and two hospitals near the center of town. Framingham residents were considered an excellent population for a long-term epidemiologic study because (1) the town's population was stable, (2) the investigators could identify a sufficient number of people with and without risk factors for heart disease (for example, they expected sufficient numbers of smokers and nonsmokers), and (3) local medical doctors were "highly cooperative and well-informed" and wanted to help investigators recruit subjects for the study. All of these factors have contributed to the study's low dropout rate (only 5% after 50 years) and to its numerous important contributions to our understanding of the etiology of heart disease.

### Selection of Comparison Group

In a cohort study, the ideal comparison group would consist of exactly the same individuals in the exposed group had they not been exposed. This theoretical concept is known as the "counterfactual ideal."<sup>19(pp49-50)</sup> Because it is impossible for the same person to be exposed and unexposed simultaneously, epidemiologists must select different sets of individuals for the exposed and comparison groups. Even though different people form each group, the investigator chooses groups as similar as possible on characteristics that influence getting the disease. For example, if gender were related to the risk of the disease (that is, if men had a higher risk than women or vice versa), it would be important for the exposed and unexposed groups to have the same proportion of males and females. The rate of disease in the exposed and comparison groups should be identical if the exposure has no effect on disease occurrence.

The three sources for the comparison group in a cohort study are (1) an internal comparison, (2) the general population, and (3) a comparison cohort (see Table 8-3). An internal comparison group is comprised of unexposed members of the same cohort. For example, Boice and Monson conducted a retrospective cohort study to determine the risk of breast cancer among female tuberculosis patients who had undergone repeated fluoroscopic x-ray examination of the chest during air collapse therapy.<sup>20</sup> This treatment resulted in considerable radiation exposure to the breast. The comparison group consisted of other female tuberculosis patients who

**TABLE 8-3** Types of Comparison Groups in Cohort Studies

Type of comparison group	Strengths	Weaknesses
Internal	Most comparable to exposed group	May be difficult to identify
General population	Accessible, stable data	Lack of comparability with exposed group; results may suffer from healthy worker effect; data on key variables may be missing
Comparison cohort	Fairly comparable	Results are often difficult to interpret because comparison cohort often has other exposures

received treatments that did not require fluoroscopic x-ray examinations. Like the exposed group, the women who comprised the internal comparison group were tuberculosis patients at Massachusetts hospitals during the years 1930 to 1954. The study found a 1.8-fold excess of breast cancer cases among the exposed group; however, the risk was higher among women exposed before age 20 years.

When the general population is the source of the comparison group, preexisting available data on disease occurrence and death in a population are the basis for comparison. In the United States, such data are available from the National Center for Health Statistics. Internationally, excellent mortality statistics are compiled by the World Health Organization. Often when the comparison group is composed of the general population, the standardized mortality ratio (SMR) is used as the measure of association. The SMR is the ratio of the observed number of cases of disease or death to the expected number based on general population rates. SMRs usually "standardize" or control for age, gender, race, and calendar time, and are interpreted like relative risks.

For example, the retrospective cohort study described earlier among workers involved in manufacturing or spraying herbicides compared the cancer mortality experience of 21,863 male and female workers to national mortality data while controlling for sex, age, and calendar time.<sup>15</sup> Because the workers came from 12 countries, mortality data collected by the World Health Organization were used to compute the expected number of deaths. Investigators found 1083 deaths due to cancer among the male workers, while 1012 were expected. These findings resulted in an SMR of 1.07. No increase in cancer mortality rates was seen in female workers.

The general population is used as a comparison group when it is not possible to find a comparable internal comparison. This problem is quite

common in occupational studies. For example, it is possible that everyone who works in a factory that manufactures herbicides is exposed to a greater or lesser degree. Although office workers, sales representatives, and company officials may be the least exposed, they usually do not comprise a comparable comparison group for production workers because of differences in gender (office workers are mainly female) and socioeconomic status (sales representatives and company officials have higher incomes).

General population comparison groups are commonly used in occupational studies of mortality because these data are quite accessible to researchers. In addition, the data are stable because they are based on large numbers. The assumption that the general population is unexposed is reasonable, because the number of exposed people is generally a small proportion of the total population.

Although there are several advantages to using the general population as a comparison group, there is one main disadvantage. The results are influenced by a form of bias known as the "healthy worker effect." This means that the rates of death and disease among a working population are usually lower than those of the general population. The lower rates of death and disease occur even when a noxious substance is present that elevates certain disease rates. The healthy worker effect occurs because a relatively healthy working population is being compared to the general population, which consists of ill people as well as healthy people. It is well known that there is selective entry of healthy persons and early removal of unhealthy ones from the workforce. In the study of workers exposed to phenoxy herbicides and chlorophenols, subjects had slightly lower than expected overall mortality rates (SMR = 0.97) that stemmed mainly from fewer deaths from circulatory system diseases (SMR = 0.91), and respiratory diseases (SMR = 0.82).<sup>15</sup>

The third type of comparison group, called the comparison cohort, consists of an unexposed cohort from another population. For example, a study compared cardiovascular disease rates among rayon factory workers who were exposed to carbon disulfide with those of paper mill workers who were not exposed to carbon disulfide.<sup>21</sup> Because both groups consisted of blue-collar workers, individual differences in social class and the healthy worker effect were minimized. However, the results of such studies are difficult to interpret because other exposures in the comparison cohort may also influence disease risk. For example, the paper mill workers may be exposed to other chemicals that influence their risk of developing cardiovascular disease.

As an alternative to a single comparison cohort, the Occupational Epidemiology Branch of the National Cancer Institute developed a comparison cohort by pooling data from earlier studies of numerous occupa-

tional populations. The assumption is that the population pooled from a variety of occupations and industries will, like the general population, have an average risk of developing particular diseases and that the healthy worker effect will be minimized.<sup>22</sup>

Epidemiologists have found that the internal comparison group is the best type of comparison group to use in a cohort study. If an internal comparison is not available, the general population is the next best option, and the traditional comparison cohort is the option of last resort. The pooled comparison cohort from the National Cancer Institute may change this ranking in the future.

## **Sources of Information**

### **Sources of Information on Exposures and Other Key Variables**

Depending on the hypothesis being tested, cohort study investigators rely on various sources for information on exposures and other important variables. These sources include medical and employment records, interviews, direct physical examinations, laboratory tests, biological specimens, and environmental monitoring. Some of these sources are preexisting, and others are designed specifically for a study. Each has advantages and disadvantages, and so investigators often use several sources to piece together all of the necessary information.

Investigators typically use health care records to describe a participant's exposure history in cohort studies of possible adverse health effects stemming from medical procedures. For example, in Boice and Monson's study of fluoroscopic x-ray exposure during tuberculosis treatment and the subsequent risk of breast cancer, investigators reviewed medical records to identify female patients who had been treated with air collapse and other therapies from 1930 through 1954.<sup>20</sup> The researchers obtained information from the medical record on the stage of the tuberculosis; all treatments received including the number of fluoroscopic x-ray examinations; and key variables such as age, race, religion, and other medical conditions.

The advantages of using medical records as an information source include low expense and a high level of accuracy and detail regarding a disease and its treatment. The medical record may be the best source of information about detailed medical events, particularly those that occurred long ago. For example, it is unlikely that a woman in Boice and Monson's study could recall the number of fluoroscopic x-ray exams that she received 30 years ago.<sup>20</sup> Her medical record could provide this important detail to the investigators to allow them to estimate dose of radiation that the woman received. The main disadvantage of medical records is that

information on important variables, apart from basic demographic characteristics, is often missing. For example, Boice and Monson had to send questionnaires to participants to obtain information on breast cancer risk factors such as family history of the disease and age at first birth, menarche, and menopause.

Employment records are used mainly to identify individuals for studies of occupational exposures. Standard data in employment records include job title, department of work, years of employment, and basic demographic characteristics such as gender and date of birth. Like medical records, employment records usually lack data on exposure details and key variables, and so investigators often have to augment them with other sources. For example, in the study of workers exposed to phenoxy herbicides and other chemicals, investigators identified cohort members not only by reviewing individual job records, but also by sending questionnaires to companies about the work environment and types of products manufactured, and by measuring chemical levels in workers' blood and fat tissue.<sup>15</sup> Even with all these sources, the study investigators acknowledged that the exposure assessment was imperfect and that some exposed individuals may have been erroneously classified as unexposed and vice versa. Furthermore, none of the sources had information on variables such as smoking, alcohol use, and other workplace exposures such as asbestos.

Because existing records have their limitations, many studies are based solely on data collected specifically for the investigation. Data collection sources include questionnaires, physical examinations, laboratory tests, biological specimens, and environmental monitoring. For example, Framingham Study participants have undergone interviews, physical exams, laboratory tests, and other tests every 2 years for the last 50 years.<sup>18</sup> The interviews have gathered information on each person's medical history, cigarette smoking, alcohol use, physical activity, dietary intake, and emotional stress. The physical exam and laboratory tests have measured, among other things, height and weight; blood pressure; vital signs and symptoms; and cholesterol, hemoglobin, and glucose levels. The biennial exams and interviews allow investigators to update changes in each person's habits and health.

Study interviews are most often administered by trained personnel either in person or over the telephone. Sometimes self-administered questionnaires are used. Interviews and questionnaires are particularly useful for obtaining information on lifestyle characteristics that are not routinely or consistently recorded in medical and other records (such as alcohol intake, sexual activity, and physical activity). Although these data tend to be better than existing records, questionnaires and interviews must still rely on the ability of participants to know and recall information. Certainly,

participants have the ability to recall important characteristics and events (such as history of cigarette smoking or family history of breast cancer), but it may be difficult for them to recall other types of information. For example, many individuals in occupational settings are not aware of the precise chemicals to which they are exposed at their job sites.

Epidemiologic studies of environmental pollution typically must conduct environmental monitoring to determine a participant's exposure. For example, investigators in the Six Cities Study, a prospective cohort study of the effect of air pollution on mortality, collected data on outdoor air concentrations of total suspended particulate matter, sulfur dioxide, ozone, and suspended sulfates from monitoring stations centrally located in the communities where the participants resided.<sup>23</sup>

When possible and appropriate, epidemiologists also gather exposure data using biological measurements in participants' blood, urine, bone, and toenails. These measurements are known as biomarkers of exposure. These markers can not only provide measures of the exposure but also provide markers of internal organ dose. For example, a prospective cohort study of the health effects of lead exposure tested children's blood lead levels at birth and several times thereafter.<sup>7</sup> The blood lead data provided an integrated assessment of each child's exposure over the previous 6 weeks. In another cohort study, investigators collected toenail clippings from study subjects in order to obtain an integrated measure of selenium intake over the previous 26 to 52 weeks.<sup>24</sup>

Data collected specifically for a study are of higher quality than those in existing records, but they are more expensive to obtain. For example, a 30-minute telephone interview used in a 1998 study to collect information on environmental exposures and confounders cost approximately \$100 to \$200 per participant.<sup>25</sup> In-person interviews, which were conducted when the subject was too ill to be interviewed by phone, were even more expensive because they involved travel time to the participant's home. Laboratory tests of biological and environmental samples can also be quite expensive. For example, analysis of polychlorinated biphenyls, dioxin, and metabolites of DDT in serum can cost about a thousand dollars per sample.

### Sources of Outcome Information

The sources of outcome information in a cohort study are similar to the sources of exposures information: interviews, self-administered questionnaires, physical examinations, biological specimens, laboratory tests, and medical records. Although accuracy is the primary consideration when selecting a source of outcome data, cost and practical considerations are also important. Investigators often rely on several sources to

gather the information because of the advantages and disadvantages described earlier.

For example, the Framingham Study used a combination of interviews, physical examinations, and laboratory tests to gather information on the incidence and precursors of coronary heart disease, stroke, and peripheral artery disease.<sup>18</sup> Medical records and laboratory tests were used to confirm information that was reported in the interview. For example, heart attacks reported by Framingham Study participants were confirmed by electrocardiographic (ECG) changes and enzyme changes diagnostic of the disorder. The presence of Q waves not previously seen was considered definitive evidence that a heart attack had occurred. The ECG evidence was useful in improving the accuracy of the heart attack data; investigators excluded about 20% of heart attacks reported by participants because of equivocal evidence.<sup>18</sup>

Biological specimens provided by the participant can also provide early evidence of disease. A biological marker is often considered an intermediate outcome because it is on the pathway from exposure to the appearance of clinical disease. For example, CD4 lymphocyte counts are markers of HIV-1 infection and AIDS disease progression. A rapid decline in CD4 counts occurs shortly after infection and continues over time through the onset of clinical disease to death.

Outcome information can be gathered from state and national disease registries (most commonly for cancer occurrence) and departments of vital records (for mortality information). Investigators can collect this type of data without directly involving the participant, which reduces study costs. In fact, mortality data are relatively easy to obtain now that the National Death Index (NDI) is available for medical research. Compiled by the U.S. National Center for Health Statistics, the NDI is a computerized central file of death record information beginning with deaths in 1979. Investigators submit identifying information to the NDI staff, who search their database for a match. The more information submitted, the greater the likelihood that an accurate match will be found. However, although mortality data are readily accessible and complete, their accuracy is often questioned because there are no strict criteria to guide physicians who are completing cause-of-death information on death certificates.

Whatever the source, it is important to follow comparable procedures for determining the outcomes among the exposed and unexposed groups. Investigators' use of differing sources and procedures may lead to inaccurate results. Thus, all resources used for one group must be used for the other. In addition, investigators should be masked to the exposure status of subjects so that they make unbiased decisions when assessing the outcomes. For example, an investigator might have to decide if an interview report of a "breast tumor" is a malignant or benign tumor. Masking

the investigator will ensure that the decision, whatever it is, is unrelated to the participant's exposure status. In addition, investigators should develop standard outcome definitions to guarantee accuracy and comparability. These definitions should be based on information that is readily available from acceptable and accessible procedures. Furthermore, diagnostic criteria should be both sensitive (able to pick up cases of disease) and specific (able to exclude individuals without disease). For example, the diagnosis of angina pectoris might include clinical symptoms such as dull, heavy discomfort in the substernal area, particularly after a meal and with exertion or emotional distress, and evidence of ECG abnormalities at rest and with exercise.

## **Approaches to Follow-up**

Loss to follow-up occurs either when a participant no longer wishes to take part in a study or when he or she cannot be located. Minimizing these losses is crucial for two reasons. First, losses to follow-up effectively decrease the sample size and reduce the ability of the study to detect an association, if one is present. Second, those who are lost to follow-up may differ in important ways from those who are successfully traced. Of particular concern is that lost individuals are more likely to have developed the disease under study. If lost individuals are more or less likely to be exposed than those successfully traced, the study results may suffer from a form of bias known as information bias (see Chapter 10 for more details).

Because high rates of follow-up are critical to the success of a cohort study, investigators have developed a variety of methods to maximize retention and trace study members.<sup>26</sup> For prospective cohort studies, strategies include collection of baseline information that helps to locate participants as the study progresses. This information usually includes the participant's full name; current address; telephone number; Social Security number; date and place of birth; and the names, addresses, and phone numbers of physicians, friends, and relatives who will always know where the participant is living.

Another recommended strategy for prospective studies is regular mail or personal contact with participants. These contacts might involve requests for up-to-date outcome information or newsletters describing the study's progress and findings. The mailing is used to help participants "bond" with the study and to obtain address corrections from the U.S. Postal Service.<sup>26</sup>

Additional mailings are the best strategy to use when participants do not initially respond. Investigators typically send at least two follow-up letters or postcards by first class or certified mail, emphasizing the

importance of the request. If additional mailings are unsuccessful, investigators use telephone calls and even home visits to contact nonrespondents.

When participants are truly lost to follow-up, researchers use a number of additional strategies.<sup>26</sup> Usually the simplest and least expensive ones are used first and more difficult and expensive ones are used later. In the early stages of follow-up, letters are sent to the last known address with "Address Correction Requested." In addition, investigators check telephone directories, directory assistance, and Internet resources such as the "White Pages." An increasingly popular resource is the U.S. Postal Service's National Change of Address (NCOA) system, which has up-to-date change-of-address data for almost the entire country.

If these steps prove unsuccessful, investigators contact relatives, friends, or physicians who were identified at baseline. If this step is unsuccessful, investigators can turn to many local, state, and national resources including state vital statistics records (for births, marriages, divorces, and deaths); driver's license, voter registration, and public utility records; the National Death Index; the Social Security Administration; the Centers for Medicare and Medicaid Services; credit bureaus; and commercial companies that specialize in tracing individuals.

Tracing is a laborious and challenging process. It has become even more difficult today because of increased population mobility. However, newly available resources such as those on the Internet offset these difficulties to some degree. The time and monetary investment in follow-up is worthwhile because it helps secure the success of the study.

## Analysis of Cohort Studies

The primary objective of the analysis of cohort study data is to compare disease occurrence in the exposed and unexposed groups. Disease occurrence is usually measured using cumulative incidence or incidence rates, and the relationship between exposure and disease occurrence is quantified by the cumulative incidence or incidence rate difference and/or ratio. Although the measures of disease frequency and association are covered in detail in Chapters 2 and 3, the calculation of person-time and the concepts of induction and latent periods are discussed here in the context of cohort study analysis.

### Calculating Person-Time

Calculating an incidence rate involves determining the amount of person-time accrued by each study subject. It is important to remember that person-time is, in essence, follow-up time and that follow-up time is calculated only within the context of a study. It is also important to differen-

tiate years of follow-up from years of exposure and years of latency. For example, consider a hypothetical occupational retrospective cohort study in which follow-up starts on January 1, 1950, and ends at the time of death or the closing date of the study (January 1, 1995). Let us consider a hypothetical person who was born on January 1, 1920; started work at age 20 on January 1, 1940; worked for 45 years until 1985; and died at age 70 on January 1, 1990 (see Figure 8-2).

The maximum number of person-years that any individual can accrue in this study is 45 years (C to F). The number of person-years that our hypothetical individual accrued was 40 years (C to E). The total number of person-years that an individual accrues never changes. However, it is often divided among various exposure categories such as age and calendar year when the individual started work, and duration of employment. Note that the number of years that our hypothetical individual worked (B to D) is a measure of exposure duration and is different from the number of person-years that he accrued in the study.

As a real-life example, let us consider Boice and Monson's study of breast cancer after repeated chest fluoroscopies.<sup>20</sup> In this study, years of follow-up were accumulated from the date of the first fluoroscopic examination for exposed women and first sanatorium admission for comparison subjects. Follow-up ended at different points, depending on what happened to the women. For women who developed breast cancer, follow-up ended with the diagnosis date. For women who did not develop breast cancer, it ended with the date of death for those who died, the closing date of the study (July 1, 1975) for those who were still alive at the end of the study, and the date last known to be alive for those who were lost to follow-up. Using these start and end dates, the 1047 women in the exposed group accrued 28,011 person-years of follow-up, and the 717 women in the comparison group accrued 19,025 person-years of follow-up. The maximum possible length of follow-up for any participant was 45 years (from first possible treatment in 1930 to closing date of study in 1975); the average was about 26 years.

	A	B	C	D	E	F
Year	1920	1940	1950	1985	1990	1995
Age (years)	0	20	30	65	70	—
Milestone	Born	Starts work	Follow-up starts	Retires	Dies	Follow-up ends

**FIGURE 8-2.** Accrual of Person-time

### Induction and Latent Periods

The analysis of a cohort study typically considers the length of time between the causal action of an exposure and the eventual diagnosis of disease.<sup>27</sup> The causal action of an exposure occurs when sufficient exposure has accrued. The *induction period* that follows is the interval between the action of a cause and disease onset. An example of disease onset might be the transformation of a normal breast cell into a cancerous one. The *latent period* is the subsequent interval between disease onset and clinical diagnosis. For example, a clinical diagnosis of breast cancer occurs when the tumor becomes large enough to be detected either by screening mammography or physical examination.

Because the time of disease onset is usually not possible to determine, the induction and latent periods are typically merged into the *empirical latent period*. For example, in a study of the relationship between tetrachloroethylene-contaminated drinking water and the risk of breast cancer, investigators hypothesized that the water contaminant might act either as a tumor initiator (early in the carcinogenic process) or as a promoter (late in the process), and so they conducted analyses considering empirical latent periods ranging from 0 to 15 years.<sup>25</sup>

### Special Types of Cohort Studies

The comparison group in a cohort study may come from either an internal source, another cohort, or the general population. Two special types of cohort studies—proportional mortality ratio (PMR) and standardized mortality ratio (SMR) studies—are conducted to compare the mortality experience of the exposed group to that of the general population. Both types of studies are common in occupational epidemiology because there is often no truly unexposed group in the workplace.

The difference between PMR and SMR studies is the type of information needed to calculate the measure of association. All of the information for the PMR study comes from death certificates (date, age, cause of death). The observed number of deaths due to a specific cause in the exposed group is compared to the expected number which is derived from the proportion of deaths resulting from that cause in the general population. The resulting PMR, which is interpreted like a relative risk, shows the relative importance of a specific cause of death in relation to all deaths.

For example, investigators compared the mortality patterns of pulp and paper mill workers to those of the U.S. population in a PMR study.<sup>28</sup> The study population consisted of members of the United Paper Workers International Union who died from 1970 through 1984, and were employed for at least 10 years in the industry. Information on year of birth, sex, race, year of death, and cause of death was abstracted from death cer-

tificates. For each cause of death, investigators compared the observed number of deaths to the expected number while controlling for sex, race, age, and calendar year of death. The study found a 31% increased risk of deaths from malignant neoplasms, particularly cancer of the lung, lymphopoietic system, and large intestine. A reduced risk of death was observed for cancers of the stomach and pancreas. The largest single category of death, arteriosclerotic heart disease, was neither increased nor reduced.

In an SMR study, investigators need additional information to compare the mortality experience of the exposed group with that of the general population. This is because the SMR compares the mortality rates of the two groups, while the PMR study compares mortality proportions. Thus, information on person-years of follow-up among the study group is needed to calculate the expected number of deaths.

For example, following publication of several PMR studies reporting increased cancer risks, investigators undertook an industry-wide SMR study of pulp and paper workers.<sup>5</sup> The study included 63,025 long-term workers from 51 mills across the United States. The workers were required to be employed for at least 10 years to be eligible. Vital status was identified through the mills, Social Security mortality tapes, and the National Death Index. Investigators calculated SMRs using three comparison populations: the U.S. population, the 20 states in which the mills were located, and the residents of the 330 counties that were within a 50-mile radius of each mill. The latter two comparison populations were used because their demographic characteristics were similar to those of the workers.

A total of 7171 deaths occurred among cohort members through the end of 1991. Person-time of follow-up was calculated for cohort members starting with 10 years after first employment or when the mill was enrolled in the study (whichever came last), and ending with the termination of the study or death. SMRs were calculated that adjusted for age, calendar time, race, and sex. The investigators found that the overall mortality rate of pulp and paper mill workers was significantly lower than that of the U.S. population (SMR=0.74) and that the workers' mortality rate was not elevated for any specific cause of death, including malignancies. However, in a few instances, a higher mortality rate was seen when workers were separated according to the type of pulping process used at their plant.

Because PMR studies take less time and cost less money, they are often done before an SMR study is initiated. However, it is informative to conduct both types of studies in the same population.<sup>29(p131)</sup> Note that SMR studies and, to a lesser extent, PMR studies suffer from the "healthy worker effect."<sup>29(p114)</sup> PMR studies also suffer from the "see-saw effect," in which deficits in one cause of death necessarily result in the study population in corresponding increases in other causes of death. This occurs

**TABLE 8-4** Strengths and Weaknesses of Cohort Studies**Strengths**

- Efficient for rare exposures
- Good information on exposures (prospective)
- Can evaluate multiple effects of an exposure
- Efficient for diseases with long induction and latent periods (retrospective)
- Less vulnerable to bias (prospective)
- Can directly measure disease incidence or risk
- Clear temporal relationship between exposure and outcome (prospective)

**Weaknesses**

- Inefficient for rare outcomes
- Poor information on exposures and other key variables (retrospective)
- Expensive and time consuming (particularly prospective)
- Inefficient for diseases with long induction and latent periods (prospective)
- More vulnerable to bias (retrospective)

because the total number of observed deaths in the study population must equal the number of expected deaths derived from the general population.

### **Strengths and Limitations of Cohort Studies**

The goal of every epidemiologic study is to harvest valid and precise information about the relationship between an exposure and a disease in a population. The various study designs merely represent different ways of harvesting this information. Each type of design has strengths and weaknesses (see Table 8-4), which tend to be complementary. For example, the pros and cons of retrospective and prospective cohort studies tend to balance one another. However, under certain circumstances, a cohort study design is clearly indicated. Investigators should use a cohort study when they wish to learn about multiple effects of an exposure or when the exposure is rare, and they should use a retrospective cohort design when the outcome of interest has a long induction and latent period.

### **Summary**

Cohort studies examine the health effects of an exposure by following two or more groups with a common characteristic (exposed and unexposed). Cohort studies may be conducted in populations that are defined by changeable conditions (open or dynamic) or irrevocable events (fixed and closed). The timing of cohort studies can be prospective, retrospective, or ambidirectional. In prospective cohort studies, investigators group participants on the basis of past or current exposure and follow them into

the future in order to observe the outcome(s) of interest. In retrospective cohort studies, investigators group participants on the basis of past exposures and evaluate outcomes that have already occurred. Ambidirectional cohort studies have retrospective and prospective components. Retrospective cohort studies are more efficient than prospective ones for studying diseases with long induction and latent periods. However, prospective studies are less vulnerable to bias than retrospective cohort studies.

Several design features of cohort studies emulate those of experimental studies in order to produce high-quality results, including the selection of comparable groups and the masking of investigators to subjects' exposure status during follow-up and outcome ascertainment. Regarding the former, the ideal comparison group would consist of the same individuals in the exposed group had they not been exposed. This is known as the "counterfactual ideal." Because it is impossible for the same individuals to be exposed and unexposed simultaneously, the investigator must select different sets of individuals for comparison. The comparison group can come from an internal comparison group, the general population, or a comparison cohort. An internal comparison group consists of unexposed members of the same cohort. A general population comparison group is selected on the basis of preexisting population data such as mortality rates from the National Center for Health Statistics. A comparison cohort consists of unexposed members of another cohort. The internal comparison group is considered the best of the three because it is most similar to the exposed group.

Depending on the hypothesis being tested, investigators rely on a variety of sources for information on exposures, outcomes, and other key variables. These sources include medical and employment records, interviews, direct physical examinations, laboratory tests, and environmental monitoring. During the follow-up period, the groups are monitored for the outcomes under study. Loss to follow-up occurs when participants no longer wish to participate or when they can no longer be located. Because follow-up losses are a threat to the validity of study, epidemiologists use a variety of methods to maximize retention and tracing of study members, including collection of baseline information to help locate participants as the study progresses.

The main strengths of cohort studies are that they are efficient for studying rare exposures, they allow investigators to evaluate multiple health effects of an exposure, and retrospective cohort studies are efficient for studying diseases with long induction and latent periods. The main weaknesses of cohort studies are that they are inefficient for studying rare outcomes, prospective cohort studies are inefficient for studying diseases with long induction and latent periods, and retrospective cohort studies are vulnerable to bias.

## References

1. Pickett JP, exec ed. *The American Heritage Dictionary of the English Language*. 4th ed. Boston, MA: Houghton Mifflin; 2000.
2. Last JM, ed. *A Dictionary of Epidemiology*. 3rd ed. New York, NY: Oxford University Press; 1995.
3. Rabkin CS, Yellin F. Cancer incidence in a population with a high prevalence of infection with human immunodeficiency virus type 1. *J Natl Cancer Inst*. 1994;86:1711-1716.
4. Kodama K, Mabuchi K, Shigematsu I. A long-term cohort study of the atomic bomb survivors. *Int J Epidemiol*. 1996;6:S95-S105.
5. Matanoski G, Kanchanaraksa S, Lees PSJ, Tao XG, Royall R, Francis M, et al. Industry-wide study of mortality of pulp and paper mill workers. *Am J Ind Med*. 1998;33:354-365.
6. Curtis KM, Savitz DA, Arbuckle TE. Effects of cigarette smoking, caffeine consumption, and alcohol intake on fecundability. *Am J Epidemiol*. 1997;146:32-41.
7. Bellinger DC, Stiles KM, Needleman HL. Low level lead exposure, intelligence and academic achievement: a long-term follow-up study. *Pediatrics*. 1992;90: 855-861.
8. Nyren O, Yin L, Josefsson S, McLaughlin JK, Blot WJ, Engqvist M, et al. Risk of connective tissue disease and related disorders among women with breast implants: a nation-wide retrospective cohort study in Sweden. *Br Med J*. 1998; 316:417-422.
9. Lathrop GD, Wolfe WH, Albanese RA, Moynahan PM. *An Epidemiologic Investigation of Health Effects in Air Force Personnel Following Exposure to Herbicides. Baseline Morbidity Results*. Washington, DC: US Air Force; 1984.
10. Wolfe WH, Michalek JE, Miner JC, Rahe A, Silva J, Thomas WF, et al. Health status of Air Force veterans occupationally exposed to herbicides in Vietnam I. Physical health. *JAMA*. 1990;264:1824-1831.
11. Michalek JE, Wolfe WH, Miner JC. Health status of Air Force veterans occupationally exposed to herbicides in Vietnam II. Mortality. *JAMA*. 1990;264: 1832-1836.
12. Vega WA, Kolody B, Hwang J, Noble A. Prevalence and magnitude of perinatal substance exposures in California. *N Engl J Med*. 1993;329:850-854.
13. Streissguth AP. Fetal alcohol syndrome: an epidemiologic perspective. *Am J Epidemiol*. 1978;107:467-478.
14. Colton T. *Statistics in Medicine*. Boston, MA: Little, Brown, and Co; 1974.
15. Kogevinas M, Becher H, Benn T, Bertazzi PA, Boffetta P, Bueno-de-Mesquita HB, et al. Cancer mortality in workers exposed to phenoxy herbicides, chlorophenols, and dioxins: an expanded and updated international cohort study. *Am J Epidemiol*. 1997;145:1061-1075.
16. Darby SC, Doll R, Gill SK, et al. Long term mortality after a single treatment course with x-rays in patients treated for ankylosing spondylitis. *Br J Cancer*. 1987;55:179-190.
17. Mills PK, Beeson WL, Phillips RL, Fraser GE. Cancer incidence among California Seventh-Day Adventists, 1976-1982. *Am J Clin Nutr*. 1994;59: 1136S-1142S.

18. Dawber TR. *The Framingham Study: The Epidemiology of Atherosclerotic Disease*. Cambridge, MA: Harvard University Press; 1980.
19. Rothman KJ, Greenland S. *Modern Epidemiology*. 2nd ed. Philadelphia, PA: Lippincott-Raven Publishers; 1998.
20. Boice JD, Monson RR. Breast cancer in women after repeated fluoroscopic examination of the chest. *J Natl Cancer Inst*. 1977;59:823-832.
21. Hernberg S, Partenen T, Nordman C-H, Sumari P. Coronary heart disease among workers exposed to carbon disulfide. *Br J Ind Med*. 1970;27:313-325.
22. Computerized Occupational Referent Population System (CORPS). Available at: <http://dceg.cancer.gov/ebp/oeb/AvailRes.html>. Accessed January 2003.
23. Dockery DW, Pope CA, Xu Spengler JD, Ware JH, Fay ME, et al. An association between air pollution and mortality in six US Cities. *N Engl J Med*. 1993; 329:1753-1760.
24. Yoshizawa K, Willett WC, Morris SJ, Stampfer MJ, Spiegelman D, Rimm EB. Study of prediagnostic selenium level in toenails and the risk of advanced prostate cancer. *J Natl Cancer Inst*. 1998;90:1219-1224.
25. Aschengrau A, Paulu C, Ozonoff D. Tetrachloroethylene-contaminated drinking water and the risk of breast cancer. *Environ Health Perspect*. 1998; 106:947-953.
26. Hunt JR, White E. Retaining and tracking cohort study members. *Epidemiol Rev*. 1998;20:57-70.
27. Rothman KJ. Induction and latent periods. *Am J Epidemiol*. 1981;114:253-259.
28. Solet D, Zoloth SR, Sullivan C, Jewett J, Michaels DM. Patterns of mortality in pulp and paper workers. *J Occup Med*. 1989;31:627-630.
29. Monson RR. *Occupational Epidemiology*. Boca Raton, FL: CRC Press; 1990.